Applications of Non-Metric Vision to Some Visual Guided Tasks

Cyril Zeller, Olivier Faugeras

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Applications of Non-Metric Vision to Some Visual Guided Tasks

Cyril ZELLER

Olivier FAUGERAS

N° 2308

Juillet 1994

PROGRAMME 4

Robotique,

image

et vision

ISSN 0249-6399

Applications of Non-Metric Vision to Some Visual Guided Tasks

Cyril ZELLER*
Olivier FAUGERAS

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n° 2308 — Juillet 1994 — 26 pages

Abstract: We present a stratification of geometric information available from stereo in three levels: Euclidean, affine and projective, depending upon the kind of calibration that has been obtained for a stereo rig. We focus on the last two levels since they are mostly unexplored. We show how projective and affine calibration can be achieved from real images without the need of calibration patterns. We also show how to use this calibration to determine, for example, whether an obstacle is coming too close to the stereo rig or such useful information as the middle of a corridor or a road.

Key-words: projective calibration, affine calibration, reconstruction, rectification

(Résumé : tsvp)

*This work was partially supported by the EEC under Esprit Project 6448, Viva

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Applications de la vision non métrique à certaines tâches visuelles

Résumé : Nous présentons une stratification de l'information géométrique disponible à partir de la stéréoscopie en trois niveaux : euclidien, affine et projectif, selon le type de calibration qui a été obtenue pour le système stéréoscopique. Nous portons surtout notre attention sur les deux derniers niveaux car ils sont beaucoup plus méconnus. Nous montrons comment une calibration projective et affine peut être obtenue à partir d'images réelles sans qu'une mire de calibration soit nécessaire. Nous montrons aussi comment utiliser cette calibration pour déterminer, par exemple, si un obstacle s'approche trop près du système stéréoscopique ou encore obtenir d'autres informations utiles telles que le milieu d'un couloir ou d'une route.

Mots-clé : calibration projective, calibration affine, reconstruction, rectification

Contents

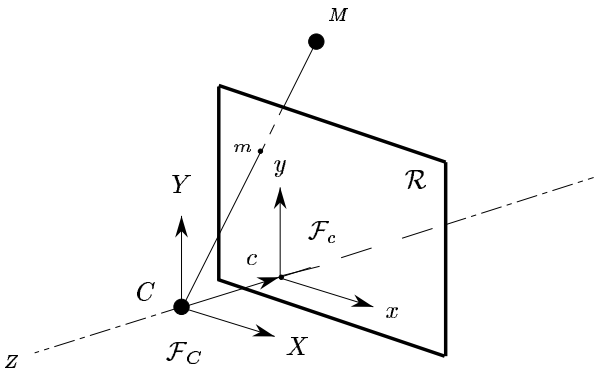

Figure 1: The pinhole model.

1 Introduction

The calibration parameters of a vision system equipping a mobile robot are likely to change over time: either because of the mechanical vibrations induced by the robot, or because some tasks to be performed by the system require to dynamically modify these parameters (tuning the zoom and focus of a lens, for instance). Now, the process of calibration used in practice is quite fastidious to be applied to the system each time needed. That is why the study of the abilities of a robot that has only a partial knowledge of the parameters of its vision system is interesting.

Section (2) describes the model used for the camera and the three-dimensional scene. After deriving from this model some relationship between two views, the consequence of a partial knowledge of it for the reconstruction of the scene is theoretically analysed. Section (3) gives some techniques to compute some of the parameters of the model without assuming full calibration of the cameras. Section (4) gives some techniques to compute information on the structure of the scene. Finally, these techniques are applied in section (5) to automatic navigation.

2 The model

2.1 The camera

The camera model used is the classical *pinhole model*. If the object space is considered to be the 3-dimensional Euclidean space $\mathcal{R}^3$ embedded in the usual way in the 3-dimensional projective space $\mathcal{P}^3$ and the image space the 2-dimensional Euclidean space $\mathcal{R}^2$ embedded in the usual way in the 2-dimensional projective space $\mathcal{P}^2$, the camera is then described as a *linear projective application* from $\mathcal{P}^3$ to $\mathcal{P}^2$ (see [1]). We can write the projection matrix

in any object frame $\mathcal{F}_O$ of $\mathcal{P}^3$:

$$\underbrace{\begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{K}} \mathbf{M}_{\mathcal{F}_O}^{\mathcal{F}_C} \tag{1}$$

where $\mathbf{A}$ is the matrix of the *intrinsic parameters*, C the optical center (see figure (1)) and $\mathbf{M}_{\mathcal{F}_i}^{\mathcal{F}_j}$ is the notation for the matrix of change of frame $\mathcal{F}_i$ to frame $\mathcal{F}_j$, such that $\mathbf{M}_{/\mathcal{F}_j} = \mathbf{M}_{\mathcal{F}_i}^{\mathcal{F}_j} \mathbf{M}_{/\mathcal{F}_i}$.

In particular, the projection equation, relating a point out of the focal plane $\mathbf{M}_{/\mathcal{F}_C}^T = [X_C, Y_C, Z_C, T_C]^T$, expressed in the normalized camera frame to its projection $\mathbf{m}^T = [x, y, 1]^T$ is

$$Z_C \mathbf{m} = \mathbf{A}\mathbf{K}\mathbf{M}_{/\mathcal{F}_C} \tag{2}$$

2.2 The scene

Even though our formalism also applies to dynamic objects, we concentrate in this paper on scenes composed of static objects. Moreover, when we study the disparity between two views, this restriction does not appear as a restriction any more if the two views have been taken simultaneously by a stereoscopic system.

2.3 The disparity between two views

We study, here, the relationship between two views of a scene. These views are supposed to come from either two cameras or one camera in motion. The optical centers corresponding to the views are denoted by C for the first and C' for the second, the intrinsic parameters by $\mathbf{A}$ and $\mathbf{A}'$ respectively, the normalized camera frames respectively by $\mathcal{F}_C$ and $\mathcal{F}_{C'}$. The matrix of change of frame $\mathcal{F}_C$ to frame $\mathcal{F}_{C'}$ is a matrix of displacement defined by a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$:

$$\mathbf{M}_{\mathcal{F}_C}^{\mathcal{F}_{C'}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \tag{3}$$

More precisely, given a point M of an object o, we are interested in establishing the disparity equation of M for the two views, that is the equation relating the projection m' of M in the second view to the projection m of M in the first view.

2.3.1 The general case

Assuming that M is not in the focal planes corresponding to the first and second views, we have, from equations (2) and (3):

$$Z'_{C'}\mathbf{m}' = \mathbf{A}'\mathbf{K}\mathbf{M}'_{/\mathcal{F}_{C'}} = \mathbf{A}' \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \end{array} \right] \mathbf{M}_{/\mathcal{F}_C} = Z_C\mathbf{A}'\mathbf{R}\mathbf{A}^{-1}\mathbf{m} + T_C\mathbf{A}'\mathbf{t}$$

We thus have obtained the general *disparity equation* relating m' to m:

$$Z'_{C'}\mathbf{m}' = Z_C\mathbf{H}_\infty\mathbf{m} + T_C\mathbf{e}' \tag{4}$$

where

$$\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1} \tag{5}$$

$$\mathbf{e}' = \mathbf{A}'\mathbf{t} \tag{6}$$

$\mathbf{H}_\infty$ is the *homography of the plane at infinity*, as detailed below in section (2.3.3). $\mathbf{e}'$ is a vector representing the *epipole* in the image frame of the second view, that is, the projection of C in the second view. Indeed, this projection is

$$\mathbf{A}'\mathbf{K}\mathbf{C}_{/\mathcal{F}_{C'}} = \mathbf{A}' \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \end{array} \right] \mathbf{C}_{/\mathcal{F}_C} = \mathbf{A}'\mathbf{t}$$

Similarly,

$$\mathbf{e} = \mathbf{A}\mathbf{R}^T\mathbf{t} \tag{7}$$

is a vector representing the epipole in the image frame of the first view.

Equation (4) means that m' lies on the line going through e' and the point represented by $\mathbf{H}_\infty\mathbf{m}$, which is the *epipolar line* of m. It is given by the vector

$$\mathbf{F}\mathbf{m} \tag{8}$$

where

$$\mathbf{F} = [\mathbf{e}']_\times\mathbf{H}_\infty \tag{9}$$

or equivalently [1],

$$\mathbf{F} = \det(\mathbf{A}')\mathbf{A}'^{-1T}[\mathbf{t}]_\times\mathbf{R}\mathbf{A}^{-1} \tag{10}$$

$\mathbf{F}$ is the *fundamental matrix* which describes the correspondence between an image point in the first view and its epipolar line in the second (see [2]).

2.3.2 The case of coplanar points

In the case of coplanar points, the equation of the plane in $\mathcal{F}_C$, relating Z_C and T_C, allows to unify their different disparity equations in one disparity equation valid for all of them.

[1] using the algebraic equation $[\mathbf{M}\mathbf{u}]_\times = \det(\mathbf{M})\mathbf{M}^{-1T}[\mathbf{u}]_\times\mathbf{M}^{-1}$, valid if $\det(\mathbf{M}) \neq 0$

The plane being given in $\mathcal{F}_C$ by the vector $\Pi^T = \begin{bmatrix} \mathbf{n}^T & -d \end{bmatrix}$, where $\mathbf{n}$ is its unitary normal in $\mathcal{F}_C$ and d, the distance from the plane to C, its equation is $\Pi^T \mathbf{M}_{/\mathcal{F}_C} = 0$, which can be written, using equation (2),

$$\begin{aligned} 0 &= \mathbf{n}^T \mathbf{K} \mathbf{M}_{/\mathcal{F}_C} - T_C d & (11) \\ &= Z_C \mathbf{n}^T \mathbf{A}^{-1} \mathbf{m} - T_C d & (12) \end{aligned}$$

If we first assume that $d \neq 0$, that is the plane does not go through C, we then obtain the new form of the disparity equation[2]:

$$Z'_{C'} \mathbf{m}' = Z_C \mathbf{H} \mathbf{m} \tag{13}$$

where

$$\mathbf{H} = \mathbf{H}_\infty + \mathbf{e}' \frac{\mathbf{n}^T}{d} \mathbf{A}^{-1} \tag{14}$$

This equation establishes the linear projective application, given by $\mathbf{H}$, the *H-matrix* of the plane, relating the projections of the points of the plane in the first view to their projections in the second. It is at the basis of the idea which consists in segmenting the scene in planar structures given by their respective H-matrices and, using this segmentation, to compute motion and structure (see [4] or [12]).

If the plane does not go either through C', its H-matrix is a homography ($\det(\mathbf{H}) \neq 0$) since its inverse is given by

$$\mathbf{H}^{-1} = \mathbf{H}' = \mathbf{H}_\infty^{-1} + \mathbf{e} \frac{\mathbf{n}'^T}{d'} \mathbf{A}'^{-1} \tag{15}$$

where $\mathbf{n}'$ is its unitary normal in $\mathcal{F}_{C'}$ and d', the distance from the plane to C'. If the plane goes through only one of the two points C or C', its H-matrix is still defined by the one of the two equations (14) or (15) which remains valid, but is no longer a homography; equation (12) shows that the plane then projects in one of the two views in a line representing by the vector

$$\mathbf{A}^{-1T} \mathbf{n} \quad \text{or} \quad \mathbf{A}'^{-1T} \mathbf{n}' \tag{16}$$

If the plane is an epipolar plane, that is goes through both C and C', its H-matrix is undefined.

Finally, equation (7) shows that e' and e always verify equation (13), as expected, since e' and e are the projections of the intersection of the line (CC') with the plane.

[2] using the algebraic equation $(\mathbf{u}^T \mathbf{M} \mathbf{v}) \mathbf{w} = (\mathbf{w} \mathbf{u}^T \mathbf{M}) \mathbf{v}$

2.3.3 The case of points at infinity

For the points of the plane at infinity, represented by $[0,0,0,1]^T$, thus of equation $T_C = 0$, the disparity equation becomes

$$Z'_{C'}\mathbf{m}' = Z_C\mathbf{H}_\infty\mathbf{m} \tag{17}$$

Thus, $\mathbf{H}_\infty$ is well the H-matrix of the plane at infinity. Equation (17) is also the limit of equation (13), when $d \to \infty$, which is compatible with the fact that the points at infinity correspond to the remote points of the scene.

2.4 The scene reconstruction

As shown by equation (2), two views are necessary to reconstruct the scene. They are also sufficient since a second equation then appears: the disparity equation studied in section (2.3).

Several kinds of reconstructions are then possible depending on the information available for the two views. Of course, the less information there is, the less precise is the reconstruction, which mathematically translates into the fact that the reconstruction is done in an Euclidean, affine or projective frame.

2.4.1 The Euclidean reconstruction

If we know $\mathbf{A}$, $\mathbf{A}'$, $\mathbf{R}$ and $\mathbf{t}$, thus $\mathbf{H}_\infty$ and $\mathbf{e}'$ through equations (5) and (6), which corresponds to a *strongly calibrated* system, then equation (4) gives us

$$\frac{Z_C}{T_C} = \frac{(\mathbf{m}' \times \mathbf{e}')(\mathbf{m}' \times \mathbf{H}_\infty\mathbf{m})}{\|\mathbf{m}' \times \mathbf{H}_\infty\mathbf{m}\|^2}$$

and equation (2)

$$\begin{bmatrix} \frac{X_C}{T_C} \\ \frac{Y_C}{T_C} \end{bmatrix} = \frac{Z_C}{T_C}\mathbf{A}^{-1}\begin{bmatrix} x \\ y \end{bmatrix}$$

M is thus reconstructed in $\mathcal{F}_C$ which corresponds to an Euclidean frame.

The projection matrices for the first and second views, expressed in their respective image frames and in $\mathcal{F}_C$, are then written

$$\mathbf{A}\begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} \quad \text{and} \quad \mathbf{A}'\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}$$

They are thus known up to an unknown displacement $\mathbf{M}_{\mathcal{F}_O}^{\mathcal{F}_C}$, where $\mathcal{F}_O$ is any Euclidean object frame.

2.4.2 The affine reconstruction

If we only know $\mathbf{H}_\infty$ and $\mathbf{F}$, thus $\mathbf{e}'$ since equation (9) shows that $\mathbf{e}'^T\mathbf{F} = 0$, both up to unknown nonzero scalar factors λ and μ,

$$\tilde{\mathbf{H}}_\infty = \lambda \mathbf{H}_\infty \quad \text{and} \quad \tilde{\mathbf{e}}' = \mu \mathbf{e}'$$

which corresponds to an *affinely calibrated* system, neither equation (2) nor equation (4) is usable since $\mathbf{A}$, $\mathbf{H}_\infty$ and $\mathbf{e}'$ are unknown. Both equations can then be rewritten in another frame $\mathcal{F}_A$ defined by its matrix of change of frame $\mathcal{F}_C$ to frame $\mathcal{F}_A$

$$\mathbf{M}_{\mathcal{F}_C}^{\mathcal{F}_A} = \begin{bmatrix} \frac{1}{\lambda}\mathbf{A} & \mathbf{0}_3 \\ \mathbf{0}_3^T & \frac{1}{\mu} \end{bmatrix}$$

If $\mathbf{M}_{/\mathcal{F}_A}^T = [X_A, Y_A, Z_A, T_A]^T$ is the vector representing M in $\mathcal{F}_A$, equation (4), written in $\mathcal{F}_A$, is

$$Z'_{C'}\mathbf{m}' = Z_A\tilde{\mathbf{H}}_\infty\mathbf{m} + T_A\tilde{\mathbf{e}}' \tag{18}$$

and equation (2),

$$Z_A\mathbf{m} = \mathbf{K}\mathbf{M}_{/\mathcal{F}_A} \tag{19}$$

Equation (18) then gives us

$$\frac{Z_A}{T_A} = \frac{(\mathbf{m}' \times \tilde{\mathbf{e}}')(\mathbf{m}' \times \tilde{\mathbf{H}}_\infty\mathbf{m})}{\|\mathbf{m}' \times \tilde{\mathbf{H}}_\infty\mathbf{m}\|^2}$$

and equation (19),

$$\begin{bmatrix} \frac{X_A}{T_A} \\ \frac{Y_A}{T_A} \end{bmatrix} = \frac{Z_A}{T_A}\begin{bmatrix} x \\ y \end{bmatrix}$$

M is thus reconstructed in $\mathcal{F}_A$ which corresponds to an affine frame.

The projection matrices for the first and second views, expressed in their respective image frames and in $\mathcal{F}_A$, are then written

$$\begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \tilde{\mathbf{H}}_\infty & \tilde{\mathbf{e}}' \end{bmatrix}$$

Indeed, the projection matrix for the second view is

$$\mathbf{A}'\mathbf{K}\mathbf{M}_{\mathcal{F}_A}^{\mathcal{F}_{C'}} = \begin{bmatrix} \mathbf{A}' & \mathbf{0}_3 \end{bmatrix}\mathbf{M}_{\mathcal{F}_C}^{\mathcal{F}_{C'}}\mathbf{M}_{\mathcal{F}_A}^{\mathcal{F}_C} = \begin{bmatrix} \mathbf{A}'\mathbf{R} & \mathbf{e}' \end{bmatrix}\begin{bmatrix} \lambda\mathbf{A}^{-1} & \mathbf{0}_3 \\ \mathbf{0}_3^T & \mu \end{bmatrix}$$

They are thus known up to the unknown affine bijection $\mathbf{M}_{\mathcal{F}_O}^{\mathcal{F}_A}$, where $\mathcal{F}_O$ is any affine object frame.

2.4.3 The projective reconstruction

If we now only know $\mathbf{F}$, up to a nonzero scalar factor, thus $\tilde{\mathbf{e}}'$, which corresponds to a *weakly calibrated* system, neither equation (2) nor equation (4) is usable since $\mathbf{A}$, $\mathbf{H}_\infty$ and $\mathbf{e}'$ are unknown. Both equations can then be rewritten in another frame $\mathcal{F}_P$ than $\mathcal{F}_C$, where the vector representing the plane at infinity is no longer known, equal to $[0,0,0,1]^T$ in $\mathcal{F}_C$ like in any Euclidean or affine frame.

For that, let us first assume that we know, up to a nonzero scalar factor λ, the H-matrix of a plane, as defined in section (2.3.2):

$$\mathbf{H} = \lambda(\mathbf{H}_\infty + \mathbf{e}'\frac{\mathbf{n}^T}{d}\mathbf{A}^{-1})$$

where $\mathbf{n}$ is the unitary normal in $\mathcal{F}_C$ of the plane and d, with $d \neq 0$, the distance from the plane to C. $\mathcal{F}_P$ is then defined by its matrix of change of frame $\mathcal{F}_C$ to frame $\mathcal{F}_P$

$$\mathbf{M}_{\mathcal{F}_C}^{\mathcal{F}_P} = \begin{bmatrix} \frac{1}{\lambda}\mathbf{A} & \mathbf{0}_3 \\ -\frac{1}{\mu}\frac{\mathbf{n}^T}{d} & \frac{1}{\mu} \end{bmatrix}$$

so that $\begin{bmatrix} \frac{\lambda}{\mu}\frac{\mathbf{n}^T}{d}\mathbf{A}^{-1} & 1 \end{bmatrix}^T$ is the vector representing the plane at infinity in $\mathcal{F}_P$. If $\mathbf{M}_{/\mathcal{F}_P}^T = [X_P, Y_P, Z_P, T_P]^T$ is the vector representing M in $\mathcal{F}_P$, we have then, using equation (2),

$$T_P = -\frac{1}{\mu}\frac{\mathbf{n}^T}{d}\mathbf{K}\mathbf{M}_{/\mathcal{F}_C} + \frac{1}{\mu}T_C \tag{20}$$

$$= -Z_C\frac{1}{\mu}\frac{\mathbf{n}^T}{d}\mathbf{A}^{-1}\mathbf{m} + \frac{1}{\mu}T_C \tag{21}$$

and, eliminating T_C from equation (4),

$$Z'_{C'}\mathbf{m}' = Z_C(\mathbf{H}_\infty + \mathbf{e}'\frac{\mathbf{n}^T}{d}\mathbf{A}^{-1})\mathbf{m} + T_P\tilde{\mathbf{e}}' \tag{22}$$

Equation (4) is thus written in $\mathcal{F}_P$

$$Z'_{C'}\mathbf{m}' = Z_P\mathbf{H}\mathbf{m} + T_P\tilde{\mathbf{e}}' \tag{23}$$

As for equation (2), it is written in $\mathcal{F}_P$

$$Z_P\mathbf{m} = \mathbf{K}\mathbf{M}_{/\mathcal{F}_P} \tag{24}$$

Equation (23) then gives us

$$\frac{Z_P}{T_P} = -\frac{(\mathbf{m}' \times \tilde{\mathbf{e}}')(\mathbf{m}' \times \mathbf{H}\mathbf{m})}{\|\mathbf{m}' \times \mathbf{H}\mathbf{m}\|^2}$$

and equation (24),

$$\begin{bmatrix} \frac{X_P}{T_P} \\ \frac{Y_P}{T_P} \end{bmatrix} = \frac{Z_P}{T_P} \begin{bmatrix} x \\ y \end{bmatrix}$$

M is thus reconstructed in $\mathcal{F}_P$ which corresponds to a projective frame. This had already been found in a quite different manner in [3] and [6].

The projection matrices for the first and second views, expressed in their respective image frames and in $\mathcal{F}_P$, are then written

$$\begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{H} & \tilde{\mathbf{e}}' \end{bmatrix}$$

Indeed, the projection matrix for the second view is

$$\mathbf{A}'\mathbf{K}\mathbf{M}_{\mathcal{F}_P}^{\mathcal{F}_{C'}} = \begin{bmatrix} \mathbf{A}' & \mathbf{0}_3 \end{bmatrix} \mathbf{M}_{\mathcal{F}_C}^{\mathcal{F}_{C'}} \mathbf{M}_{\mathcal{F}_P}^{\mathcal{F}_C} = \begin{bmatrix} \mathbf{A}'\mathbf{R} & \mathbf{e}' \end{bmatrix} \begin{bmatrix} \lambda \mathbf{A}^{-1} & \mathbf{0}_3 \\ \lambda \frac{\mathbf{n}^T}{d} \mathbf{A}^{-1} & \mu \end{bmatrix}$$

and is well of rank 3 as product of a 3×4-matrix of rank 3 and a 4×4-matrix of rank 4.

They are thus known up to the unknown homography $\mathbf{M}_{\mathcal{F}_O}^{\mathcal{F}_P}$, where $\mathcal{F}_O$ is any projective object frame.

The reconstruction described above is possible as soon as the H-matrix of a plane which does not go through C is known. In particular, when $\mathbf{F}$ is known, one is always available as equations (9) and (22) suggest it. It is defined by

$$\frac{\mathbf{n}^T}{d} = -\frac{\mathbf{e}'^T}{\|\mathbf{e}'\|^2} \mathbf{H}_\infty \mathbf{A} = -\frac{\mathbf{t}^T \mathbf{A}'^T \mathbf{A}'\mathbf{R}}{\|\mathbf{A}'\mathbf{t}\|^2} \tag{25}$$

which gives, using equation (9),[3]

$$\mathbf{H} = [\frac{\mathbf{e}'}{\|\mathbf{e}'\|}]_\times \mathbf{F}$$

The equation, expressed in $\mathcal{F}_{C'}$, of the corresponding plane is $\begin{bmatrix} \mathbf{n}^T & -d \end{bmatrix} \mathbf{M}_{\mathcal{F}_{C'}}^{\mathcal{F}_C} \mathbf{M}_{/\mathcal{F}_{C'}} = 0$, thus, using equation (25),

$$\mathbf{e}'^T \mathbf{A}'\mathbf{K}\mathbf{M}_{/\mathcal{F}_{C'}} = 0$$

which shows, using equation (2), that this plane is the plane which projects, in the second view, to the line representing by $\mathbf{e}'$, as already noticed in [9].

[3] using the algebraic equation $\mathbf{u}\mathbf{u}^T = \|\mathbf{u}\|^2 \mathbf{I}_3 + [\mathbf{u}]_\times^2$

3 Computing some model parameters

Now that we have established a model and seen which of its parameters are necessary to deduce information on the structure of the scene, we go over some methods to compute some of these parameters, giving images of real cameras.

If no a priori knowledge, nor of some parameters of the model, neither of the scene, is assumed, the only entrance left in the model is through the knowledge of point correspondences between the two images.

3.1 The correspondences

The correspondence matching is done using the image intensity function $I(x, y)$. A criterion, usually depending on the local value of $I(x, y)$, is chosen to decide whether a point of the first image and a point of the second correspond to the projections of the same point of the scene. It is generally more or less based on some physical model of the scene.

The context in which the views have been taken plays a significant role. In particular, two important cases have to be considered: the case where the views are very similar and the opposite case. The first case usually corresponds to views of a sequence taken by one camera, the second, to views taken by a stereoscopic system of two cameras. In the first case, tracking points using a simple correlation criterion yields good results. The second case requires more sophisticated criteria. This sophistication is the price to pay if we want to manipulate pairs of simultaneous shots, which allow general reconstruction of the scene without worrying about the motion of the objects observed, as underlined in section (2.2).

In both cases, the evaluation of the criterion is not performed for all the image points, but only for predetermined points of interest. These points are usually the corners of the image, given by the maxima of some operators applied to $I(x, y)$. Indeed, they are the most likely to be invariant to view changes for these operators.

The implementation of the corner detector. The operator used is the one considered in [5], which is a slightly modified version of the Plessey corner detector:

$$\det(\hat{\mathbf{C}}) - k(\mathrm{trace}(\hat{\mathbf{C}}))^2$$

where

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{I}_x^2 & \hat{I_x I_y} \\ \hat{I_x I_y} & \hat{I}_y^2 \end{bmatrix}$$

and $\hat{I}$ denotes the smoothing operation on I. Taking k equal to 0.04 and thresholding the result leads to corner detection.

The implementation of the points tracker. The implementation has been strongly influenced by the corner tracker described in [7].

The correlation criterion used is:

$$C(p_1, p_2) = \cos(i_1 - \overline{i_1}, i_2 - \overline{i_2}) = \frac{(i_1 - \overline{i_1})(i_2 - \overline{i_2})}{\|i_1 - \overline{i_1}\|\|i_2 - \overline{i_2}\|}$$

where i_j is the vector of the local image intensity function around the point p_j and $\overline{i_j}$ its mean.

The application of this criterion is explained in the sequel.

First, the corners are extracted in both images.

Then, for a given corner of the first image, the following operation is performed: its neighborhood is searched for corners of the second image; the criterion is applied to each pair formed by the corner of the first image and one of these corners found; the scores are thresholded and, if there are pairs left, the one which obtained the best score is retained as a correspondence.

After that, for each corner of the second image for which a correspondent point in the first image has been found, the preceding operation is applied from the second image to the first. If the correspondent point found by this operation is the same as the previous one, it is then definitely taken as valid.

The implementation of the stereo points matcher. The method described in the previous section no longer works as soon as the views are quite different. More precisely, the correlation criterion is not selective enough: there are, for a given point of an image, several points of the other image that lead to a good correlation score, without the best of them being the real correspondent point searched. To achieve correspondence matching, the process must then keep all those potentially good but conflicted correspondences and involves global techniques to decide between them: a classical relaxation technique is used to converge towards a globally coherent system of point correspondences, giving some constraints of uniqueness and continuity (see [13]).

3.2 The fundamental matrix

Once some image point correspondences, represented in the image frame by $(\mathbf{m}'_i, \mathbf{m}_i)$, have been found, the fundamental matrix $\mathbf{F}$ is computed, up to a nonzero scalar factor, as the unique solution of the system of equations, derived from the disparity equations,

$$\mathbf{m}_i'^T \mathbf{F} \mathbf{m}_i = 0 \tag{26}$$

This system can be solved as soon as seven such correspondences are available: only eight coefficients of $\mathbf{F}$ need to be computed, since $\mathbf{F}$ is defined up to a nonzero scalar factor, while equation (26) supplies one scalar equation per correspondence and $\det(\mathbf{F}) = 0$, the eighth. If there are more correspondences available, which are not exact, as it is the case in practice, the goal of the computation is to find the matrix which approximates at best the solution of this system by least squares according to a given criterion.

A study of the computation of the fundamental matrix from image point correspondences can be found in [8]. Here, we just mention our particular implementation, which consists, on the one hand, in a direct computation considering that all the correspondences are valid and in the other hand, in a method to reject some possible outliers among the correspondences.

The direct computation computes $\mathbf{F}$ in order to minimize the following criterion:

$$\sum_i \left(\frac{1}{[\mathbf{F}\mathbf{m}_i]_x^2 + [\mathbf{F}\mathbf{m}_i]_y^2} + \frac{1}{[\mathbf{F}^T\mathbf{m}'_i]_x^2 + [\mathbf{F}^T\mathbf{m}'_i]_y^2} \right) (\mathbf{m}_i'^T \mathbf{F}\mathbf{m}_i)^2$$

which is the sum of the squares of the distance of m_i to the epipolar line of m'_i and the distance of m'_i to the epipolar line of m_i. This minimization is performed by a classical Levenberg-Marquardt method (see [10]). In order to take in account both its definition up to a scalar product and the fact that it is of rank 2, a parametrization of $\mathbf{F}$ in seven parameters is used, which parametrizes all the 3×3-matrices of rank strictly less than 3. These parameters are computed from $\mathbf{F}$ the following way: a line l and a column c of $\mathbf{F}$ are chosen to be written as a linear combination of the other lines and other columns; the four coefficients of these two combinations are taken as parameters; among the four coefficients not belonging to l and c, the three smallest, in absolute value, are divided by the biggest and taken as the last three parameters. l and c are chosen in order to maximize the rank of the derivative of $\mathbf{F}$ with respect to the parameters. Denoting the parameters by $p1$, $p2$, $p3$, $p4$, $p5$, $p6$ and $p7$ and assuming, for instance, l and c equals to 1 and the bottom right coefficient being the normalized coefficient, lead to the following matrix:

$$\begin{bmatrix} p6(p4p1 + p5p3) + p7(p4p2 + p5) & p4p1 + p5p3 & p4p2 + p5 \\ p6p1 + p7p2 & p1 & p2 \\ p6p3 + p7 & p3 & 1 \end{bmatrix}$$

During the process of minimization, the parametrization of $\mathbf{F}$ is susceptible of change: the parametrization chosen for the matrix at the beginning of the process is not necessarily the most suitable for the final matrix.

The outliers rejection method used is a classical least median of squares method. It is described in detail in [13].

3.3 The H-matrix of a plane

If we have at our disposal correspondences, represented in the image frame by $(\mathbf{m}'_i, \mathbf{m}_i)$, of points belonging to a plane, the H-matrix $\mathbf{H}$ of this plane is computed, up to a nonzero scalar factor, as the unique solution of the system of equations (13),

$$Z'_{C'}\mathbf{m}'_i = Z_C \mathbf{H}\mathbf{m}_i$$

This system can be solved as soon as four such correspondences are available: only eight coefficients of $\mathbf{H}$ need to be computed, since $\mathbf{H}$ is defined up to a nonzero scalar factor,

while equation (13) supplies two scalar equations per correspondence. If there are more correspondences available, which are not exact, as it is the case in practice, the goal of the computation is to find the matrix which approximates at best the solution of this system according to a given criterion: a study of the computation of plane H-matrices from image point correspondences can be found in [1].

In particular, three points define a plane, whose H-matrix is computable as soon as we know e' and e (for instance, through the fundamental matrix), since e' and e verify equation (13). If the plane is defined by one point and a line L, given by its projections (l, l'), so that e does not belong to l and e' does not belong to l', its H-matrix is computable the same way, as soon as we know the fundamental matrix. Indeed, the projections of two other points M and N of the plane are given by choosing two points m and n on l, which amounts to choosing M and N on L: the corresponding points m' and n' are then given by intersecting l' with the epipolar line of m and the epipolar line of n, given by the fundamental matrix.

Given the H-matrix $\mathbf{H}$ of a plane P and the correspondences (m, m') and (n, n') of two points M and N, it is possible to directly compute in the images the correspondences (i, i') of the intersection I of the line (MN) with P. Indeed, i' belongs both to $(m'n')$ and the image of (mn) by $\mathbf{H}$, thus:

$$\mathbf{i}' = (\mathbf{m}' \times \mathbf{n}') \times (\mathbf{Hm} \times \mathbf{Hn})$$

(see [11])

Similarly, given two planes P and Q by their H-matrices $\mathbf{H}_P$ and $\mathbf{H}_Q$, it is possible to directly compute in the images the correspondences of the intersection L of P with Q. Indeed, the correspondences of two points of L are computed, for example, as intersections of two lines L_1 and L_2 of P with Q; the correspondences of such lines are obtained by choosing two lines in the first image representing by the vectors $\mathbf{l}_1$ and $\mathbf{l}_2$, their corresponding lines in the second image being given by $\mathbf{H}_P^{-1T}\mathbf{l}_1$ and $\mathbf{H}_P^{-1T}\mathbf{l}_2$.

3.4 The homography of the plane at infinity

To compute the homography of the plane at infinity $\mathbf{H}_\infty$, we can no longer use the disparity equation (4) with correspondences of points not at infinity, even if we know the fundamental matrix, since $Z'_{C'}$, Z_C and T_C are not known. We must, thus, have at our disposal correspondences of points at infinity $(\mathbf{m}'_i, \mathbf{m}_i)$ and compute $\mathbf{H}_\infty$ like any other plane H-matrices, as described in section (3.3).

The only way to obtain some correspondences of points at infinity is to assume some additional knowledge.

In a first way, we can assume that we have some additional knowledge of the observed scene that allows to identify, in the images, some projections of points at infinity, like, for instance, the intersections of the projections of parallel lines of the scene, or some projections of points at the horizon, which provide sufficiently good approximations of points at infinity.

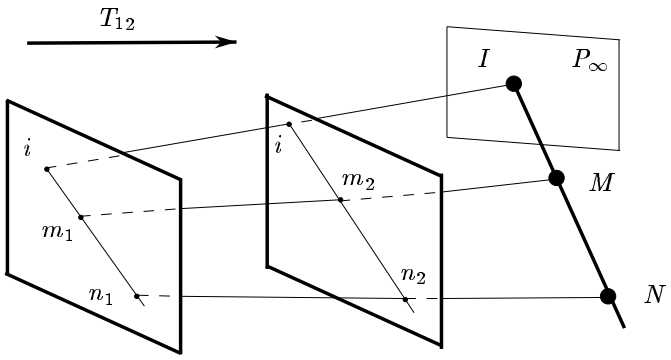

Figure 2: Determining the projections of points at infinity (see section (3.4)).

Another way to proceed is to assume that we have an additional pair of views. More precisely, if this second pair differs from the first only by a translation, any pair (M, N) of stationary object points (see figure (2)), given in the first views by $(\mathbf{m}_1, \mathbf{m}'_1)$ and $(\mathbf{n}_1, \mathbf{n}'_1)$, and, in the second, by $(\mathbf{m}_2, \mathbf{m}'_2)$ and $(\mathbf{n}_2, \mathbf{n}'_2)$, gives us the projections $(\mathbf{i}_1, \mathbf{i}'_1)$ and $(\mathbf{i}_2, \mathbf{i}'_2)$ in the four images of the intersection I of the line (MN) with the plane at infinity. Indeed, on the one hand, since I is at infinity and the stationarity of M and N implies the stationarity of I, we have, from equations (17) and (4),

$$Z_{C_2}\mathbf{i}_2 = Z_{C_1}\mathbf{H}_{\infty 12}\mathbf{i}_1 \quad \text{and} \quad Z'_{C'_2}\mathbf{i}'_2 = Z'_{C'_1}\mathbf{H}'_{\infty 12}\mathbf{i}'_1$$

and, in the case where the two pairs of views differ only by a translation, $\mathbf{A}_1 = \mathbf{A}_2$, $\mathbf{R}_{12} = \mathbf{I}_3$, $\mathbf{A}'_1 = \mathbf{A}'_2$, $\mathbf{R}'_{12} = \mathbf{I}_3$ and we have, by equation (5),

$$\mathbf{H}_{\infty 12} = \mathbf{I}_3 \quad \text{and} \quad \mathbf{H}'_{\infty 12} = \mathbf{I}_3$$

which implies that $i_1 = i_2 = i$ and $i'_1 = i'_2 = i'$. On the other hand, as I lies on (MN), i_1 lies on $(m_1 n_1)$, i_2, on $(m_2 n_2)$, i'_1, on $(m'_1 n'_1)$ and i'_2, on $(m'_2 n'_2)$. Consequently, i and i' are obtained as the intersections, respectively, of $(m_1 n_1)$ with $(m_2 n_2)$ and of $(m'_1 n'_1)$ with $(m'_2 n'_2)$:

$$\mathbf{i} = (\mathbf{m}_1 \times \mathbf{n}_1) \times (\mathbf{m}_2 \times \mathbf{n}_2) \quad \text{and} \quad \mathbf{i}' = (\mathbf{m}'_1 \times \mathbf{n}'_1) \times (\mathbf{m}'_2 \times \mathbf{n}'_2)$$

Once $\mathbf{H}_\infty$ has been obtained, any ratios of three aligned points of the scene can be directly computed in the images. Indeed, given three points M_1, M_2 and M_3 on a line, like in figure (3), by their projections (m_1, m'_1), (m_2, m'_2) and (m_3, m'_3) on the images, we have the projections (m, m') of the intersection of this line with the plane at infinity, using $\mathbf{H}_\infty$, as explained in section (3.3). We can thus compute the cross-ratio of these four projections. As projective invariant, this cross-ratio is then exactly equal to the ratio of M_1, M_2 and M_3.

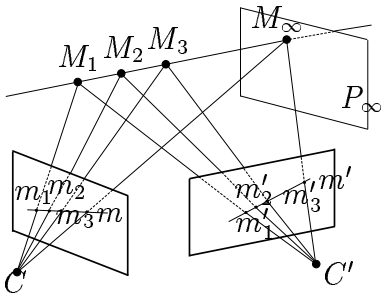

Figure 3: Determining ratios of distances in affine calibration: $\frac{\overline{M_1M_3}}{\overline{M_2M_3}} = \frac{\overline{M_1M_3}}{\overline{M_1M_\infty}} : \frac{\overline{M_2M_3}}{\overline{M_2M_\infty}} = \frac{\overline{m_1m_3}}{\overline{m_1m}} : \frac{\overline{m_2m_3}}{\overline{m_2m}}$ (see section (3.4)).

4 Positioning points with respect to a plane

In this section, we assume that we know some of the model parameters which lets us perform a *rectification with respect to a plane* of the scene. This process, explained below, allows not only to compute a map of image point correspondences, but also to assign to each of them a scalar that represents a measure of the disparity between the two projections of the correspondence. This number is in turn related to the position of the corresponding point of the scene with respect to the plane.

4.1 The rectification with respect to a plane

Like in section (2.4.3), we assume that we know, up to nonzero scalar factors, $\mathbf{F}$, thus $\tilde{\mathbf{e}}'$, and the H-matrix $\mathbf{H}$ of a plane P. Let us then choose two homographies, represented by the matrices $\hat{\mathbf{H}}'$ and $\hat{\mathbf{H}}$, such that

$$\hat{\mathbf{H}}'\tilde{\mathbf{e}}' = \alpha[1,0,0]^T \quad (27)$$

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}'\mathbf{H} \quad (28)$$

where α is any scalar. Equation (23) can then be rewritten

$$Z'_{C'}\hat{\mathbf{m}}' = Z_P\hat{\mathbf{m}} + T_P\alpha[1,0,0]^T \quad (29)$$

where

$$\hat{\mathbf{m}}' = \hat{\mathbf{H}}'\mathbf{m}' \quad \text{and} \quad \hat{\mathbf{m}} = \hat{\mathbf{H}}\mathbf{m}$$

The rectification with respect to a plane consists in applying such matrices, called the *rectification matrices*, $\hat{\mathbf{H}}'$ to the second image and $\hat{\mathbf{H}}$ to the first.

Equation (29) shows that the corresponding point $\hat{m}'$ in the second rectified image of a point $\hat{m}$ of the first rectified image then lies on the line parallel to the x-axis and going

through $\hat{m}$. Applying a correlation criterion to $\hat{m}$ and each point of this line thus allows to determine $\hat{m}'$, if the image is not too distorted through the process of rectification. Equations (27) and (28) do not completely determine $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}'$: this indetermination is used to minimize the distortion of the images, as explained in section (4.2).

Once $\hat{m}'$ has been determined, a measure of the disparity between $\hat{m}'$ and $\hat{m}$ with respect to this plane is given by the difference between the x-coordinate of $\hat{m}'$ and the x-coordinate of $\hat{m}$. If M belongs to P, it is equal to zero since T_P then vanishes as shown by equations (11) and (20); otherwise, its interpretation depends on the information available for the model, as explained in sections (4.3) and (4.4).

4.2 The minimization of the distortion

4.2.1 A decomposition of an homography

For any matrix $\mathbf{H}$, $\mathbf{H}\mathbf{H}^T$ is clearly a symmetric matrix. If, moreover, $\mathbf{H}$ is a square nonsingular matrix, then $\mathbf{H}\mathbf{H}^T$ is also positive definite. Indeed, for each vector $\mathbf{v}$, different from $\mathbf{0}$,

$$\mathbf{v}^T\mathbf{H}\mathbf{H}^T\mathbf{v} = \|\mathbf{H}^T\mathbf{v}\|^2 > 0$$

since the kernel of $\mathbf{H}^T$ reduced to $\mathbf{0}$ because of the nonsingularity of $\mathbf{H}$. Consequently, a nonsingular upper triangular matrix $\mathbf{U}$ exists such that[4]

$$\mathbf{H}\mathbf{H}^T = \mathbf{U}\mathbf{U}^T$$

If we denote by $\mathbf{O}$ the matrix $\mathbf{U}^{-1}\mathbf{H}$, we see that

$$\mathbf{O}\mathbf{O}^T = \mathbf{U}^{-1}\mathbf{H}\mathbf{H}^T\mathbf{U}^{-1T} = \mathbf{I}$$

which shows that $\mathbf{O}$ is an orthogonal matrix.

Finally, we have the following decomposition of $\mathbf{H}$:

$$\mathbf{H} = \mathbf{U}\mathbf{O} \tag{30}$$

where $\mathbf{U}$ is a nonsingular upper triangular matrix and $\mathbf{O}$, an orthogonal matrix. By decomposing $\mathbf{H}^{-1}$ like in equation (30), inverting it and knowing that the inverse of an upper triangular matrix is also an upper triangular matrix, we see that $\mathbf{H}$ can also symmetricaly be written as an orthogonal matrix right multiplied by an upper triangular matrix.

4.2.2 How many degrees of freedom are left ?

$\mathbf{H}$ being known, equation (28) shows that $\hat{\mathbf{H}}$ is completely determined as soon as $\hat{\mathbf{H}}'$ is. So, all the degrees of freedom left are concentrated in $\hat{\mathbf{H}}'$. Only eight coefficients of $\hat{\mathbf{H}}'$ need to be

[4] It is one form of the Cholesky decomposition.

computed, since $\hat{\mathbf{H}}'$ is defined up to a nonzero scalar factor, and equation (27) supplies two scalar equations: six degrees of freedom remain, but how many of them are really involved in the distortion ?

According to section (4.2.1), two matrices, $\mathbf{U}$ and $\mathbf{R}$ exist such that

$$\hat{\mathbf{H}}' = \mathbf{U}\mathbf{R}$$

$\mathbf{U}$ is an upper triangular matrix and $\mathbf{R}$, a rotation matrix. If we decompose $\mathbf{R}$ as a product of three rotations around the x- y- and z-axis, we can write

$$\hat{\mathbf{H}}' = \underbrace{\begin{bmatrix} \mathbf{U}_2 & \mathbf{v} \\ \mathbf{0}_3^T & \lambda \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \mathbf{R}_2 & \mathbf{0}_2 \\ \mathbf{0}_3^T & 1 \end{bmatrix}}_{\mathbf{R}_z} \mathbf{R}_y\mathbf{R}_x = \begin{bmatrix} \mathbf{U}_2\mathbf{R}_2 & \mathbf{v} \\ \mathbf{0}_3^T & \lambda \end{bmatrix} \mathbf{R}_y\mathbf{R}_x$$

where $\mathbf{U}_2$ is an upper triangular matrix, $\mathbf{R}_2$, a rotation matrix, $\mathbf{v}$ a vector and λ, a scalar. Now, according to section (4.2.1), $\mathbf{U}_2\mathbf{R}_2$ can be rewritten as $\mathbf{R}_2'\mathbf{U}_2'$ where $\mathbf{R}_2'$ is a rotation matrix and $\mathbf{U}_2'$, an upper triangular matrix and we can write

$$\hat{\mathbf{H}}' = \underbrace{\begin{bmatrix} \mathbf{R}_2' & \mathbf{0}_2 \\ \mathbf{0}_3^T & 1 \end{bmatrix}}_{\mathbf{R}_z'} \underbrace{\begin{bmatrix} \mathbf{U}_2' & \mathbf{R}_2'^T\mathbf{v} \\ \mathbf{0}_3^T & \lambda \end{bmatrix}}_{\mathbf{U}'} \mathbf{R}_y\mathbf{R}_x$$

where $\mathbf{R}_z'$ is a simple rotation in the plane and $\mathbf{U}'$, an upper triangular matrix. Lastly, if we extract from $\mathbf{U}'$ the translation and scaling components, we have

$$\hat{\mathbf{H}}' = \lambda\mathbf{R}_z' \begin{bmatrix} s_x & 0 & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & s_{xy} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R}_y\mathbf{R}_x \tag{31}$$

Based on equation (31), $\mathbf{R}_y$ is chosen in order to cancel out the third coordinate of $\hat{\mathbf{H}}'\tilde{\mathbf{e}}'$, involved in equation (27), (making the epipolar lines parallel) and $\mathbf{R}_z'$, in order to cancel out its second coordinate (making the epipolar lines parallel to the x-axis). Since the scaling factors, s_x and s_y, and the translation terms, u_0 and v_0, are not involved in the distortion, two degrees of freedom are left, given by s_{xy} and $\mathbf{R}_x$.

4.2.3 A criterion for the distortion

The criterion to be minimized is the ratio of the surface of the rectangle circumscribing the rectified image to the surface of the rectified image.

This criterion is valid as soon as these surfaces are not infinite, that is, as soon as the lines l and l', which are mapped, by, respectively, $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}'$, to the line at infinity, do not go through any point of, respectively, the first and second images. If e and e' do not lie in, respectively, the first and second image, $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}'$ can be chosen to verify this constraint,

since equations (27) and (28) show that l and l', which are represented by the last rows of, respectively, $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}'$, are only constrained to go through, respectively, e and e'.

4.3 The interpretation of the disparity in the Euclidean case

If we know $\mathbf{A}$, $\mathbf{A}'$, $\mathbf{R}$ and $\mathbf{t}$, the most interesting result is given by a rectification with respect to the plane at infinity. In that case, equation (29) is written

$$Z'_{C'}\hat{\mathbf{m}}' = Z_C\hat{\mathbf{m}} + T_C\alpha[1,0,0]^T$$

Then choosing

$$\hat{\mathbf{H}}' = \mathbf{B}\hat{\mathbf{R}}'\mathbf{A}'^{-1} \quad \text{and} \quad \alpha = 1 \tag{32}$$

where $\hat{\mathbf{R}}'$ is a rotation matrix such that

$$\hat{\mathbf{R}}'\mathbf{t} = [1,0,0]^T \tag{33}$$

and $\mathbf{B}$ is any matrix of intrinsic parameters, leads to

$$\hat{Z}'_{C'}\hat{\mathbf{m}}' = \hat{Z}_C\hat{\mathbf{m}} + \hat{T}_C\beta_u[1,0,0]^T \tag{34}$$

where

$$\hat{Z}'_{C'}\hat{\mathbf{m}}' = Z'_{C'}\mathbf{B}\hat{\mathbf{R}}'\mathbf{A}'^{-1}\mathbf{m}' \quad \text{and} \quad \hat{Z}_C\hat{\mathbf{m}} = Z_C\mathbf{B}\hat{\mathbf{R}}'\mathbf{R}\mathbf{A}^{-1}\mathbf{m} \tag{35}$$

and β_u is the top left element of $\mathbf{B}$. If we define the two rectified frames $\hat{\mathcal{F}}_C$ and $\hat{\mathcal{F}}_{C'}$ such that

$$\mathbf{M}_{\mathcal{F}_{C'}}^{\hat{\mathcal{F}}_{C'}} = \begin{bmatrix} \hat{\mathbf{R}}' & \mathbf{0}_3 \\ \mathbf{0}_3^T & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{M}_{\mathcal{F}_C}^{\hat{\mathcal{F}}_C} = \begin{bmatrix} \hat{\mathbf{R}}'\mathbf{R} & \mathbf{0}_3 \\ \mathbf{0}_3^T & 1 \end{bmatrix} \tag{36}$$

equations (35) are interpreted as the disparity equations respectively, of the views corresponding to $\mathcal{F}_C$ and $\hat{\mathcal{F}}_C$ and the views corresponding to $\mathcal{F}_{C'}$ and $\hat{\mathcal{F}}_{C'}$. $\mathbf{B}$ then corresponds to the intrinsic parameters of the rectified views. Equation (32) lets two parameters from $\mathbf{B}$ and one from $\hat{\mathbf{R}}'$, so three parameters to minimize the distortion. Indeed, among the five parameters of $\mathbf{B}$, the two parameters of the translation and the scaling factor are not involved in the distortion and equation (33) shows that $\hat{\mathbf{R}}'$ is defined up to an unknown rotation around $\mathbf{t}$.

From equation (36), we have

$$\mathbf{M}_{\hat{\mathcal{F}}_C}^{\hat{\mathcal{F}}_{C'}} = \mathbf{M}_{\mathcal{F}_{C'}}^{\hat{\mathcal{F}}_{C'}}\mathbf{M}_{\mathcal{F}_C}^{\mathcal{F}_{C'}}\mathbf{M}_{\hat{\mathcal{F}}_C}^{\mathcal{F}_C} = \begin{bmatrix} \mathbf{I}_3 & [1,0,0]^T \\ \mathbf{0}_3^T & 1 \end{bmatrix}$$

which shows that $\hat{Z}'_{C'} = \hat{Z}_C$. A measure of the disparity is then given by

$$\frac{\hat{T}_C}{\hat{Z}_C}\beta_u$$

and thus is proportional to the inverse of the distance of M to the focal plane of the rectified images.

4.4 The interpretation of the disparity in the affine and projective case

If we know only $\mathbf{F}$, up to a nonzero scalar factor, we do not have the rectification matrices of the Euclidean case. Representing $\hat{m}'$ by $[\hat{x}', \hat{y}', \hat{z}']^T$, $\hat{m}$ by $[\hat{x}, \hat{y}, \hat{z}]^T$ and using equation (29), (20) and (3), a measure $\mathcal{D}$ of the disparity is then given by

$$\mathcal{D} = \frac{\hat{x}'}{\hat{z}'} - \frac{\hat{x}}{\hat{z}} = \frac{T_P}{Z'_{C'}} \frac{\alpha}{\hat{z}'} = \frac{1}{\mu d} \frac{d(M,P)}{d(M,P'_f)} \frac{\alpha}{k d(m',l')}$$

where

$$d(M,P) = d - \mathbf{n}^T [\frac{X_C}{T_C}, \frac{Y_C}{T_C}, \frac{Z_C}{T_C}]^T \quad \text{and} \quad d(M,P'_f) = \frac{Z'_{C'}}{T'_{C'}} \tag{37}$$

$$d(m',l') = \frac{\mathbf{l}'^T \mathbf{m}'}{k} = \frac{\hat{z}'}{k} \quad \text{and} \quad k = \sqrt{r'^2_{31} + r'^2_{32}}$$

and $\mathbf{l}'^T = [r'_{31}, r'_{32}, r'_{33}]^T$ denotes the last row of $\hat{\mathbf{H}}'$. $d(M,P)$, respectively $d(m,l)$, is such that its absolute value is the Euclidean distance of M, respectively m, to P, respectively l, and its sign, the same for all the points located on the same side of P, respectively l.

The sign of $\mathcal{D}$ thus tells us the position of M with respect to P and the focal plane P'_f corresponding to the second view. Indeed, μ, d, α and k do not depend on M and we can choose $\hat{\mathbf{H}}'$ such that l' does not go through any point of the second image, as described above, so that the sign of $d(m',l')$ does not depend on m'. But, the interesting measure is, above all, $\hat{z}'\mathcal{D}$, whose absolute value is moreover proportional to the ratio of the distance of M to P to the distance of M to P'_f (see also [12]).

If $\mathbf{H}_\infty$ is also known, the rectification can be done with respect to the plane at infinity and gives the following disparity measure:

$$\mathcal{D}_\infty = \frac{\alpha}{\mu d(M,P'_f) k d(m',l)} \tag{38}$$

As expected, $\hat{z}'\mathcal{D}_\infty$ is of constant sign for all the points that are on the same side of P'_f and its absolute value is inversely proportional to the distance of M to P'_f.

5 An application to robot navigation

This section describes how to give sight to a robot, equipped with a stereoscopic system of two cameras, using a very simple method to calibrate them. Both, the projective case and the affine case are considered.

5.1 The projective calibration

Let us imagine the following calibration stage:

- as described in section (3.1), some correspondences between two views taken by the cameras are found;
- these correspondences are used to compute the fundamental matrix, as described in section (3.2);
- three particular correspondences are given to the system; they correspond to three object points defining a virtual plane P in front of the robot, at known distance d_0;
- the H-matrix of P is computed as described in section (3.3);

The fundamental matrix, as well as the plane H-matrix, remain the same for any other pair of views taken by the system, as long as the intrinsic parameters of both cameras and the attitude of one camera with respect to the other do not change.

According to section (4), by indefinitely performing rectifications with respect to P, the robot then permanently knows whether there are points in front of itself up to the distance d_0, by looking at the sign of their disparity and can act in consequence: either avoiding the points, or following them. Furthermore, if P and P'_f intersect sufficiently far away, it can detect whether the points are moving away or towards itself. Indeed, equation (37) shows that, for each point M, the absolute value of $\hat{z}'\mathcal{D}$ is then approximatively proportional to

$$|1 - \frac{d_0}{|d(M, P'_f)|}|$$

thus, is a decreasing function of the distance of M to the focal plane of one camera.

At last, since we are only interested in the points around the plane, which have a null disparity, we can limit the search along the epipolar line of the correspondent point $\hat{m}'$ of any point $\hat{m}$ to an interval around $\hat{m}$, which significantly reduces the computation time.

An example is given in figures (4), (5), (6) and (7). Figure (4) shows as dark square boxes the points used to define a plane and the image of a fist taken by the left camera. Figure (5) shows the left and right images once rectified with respect to this plane. Figure (6) shows the disparity map obtained by correlation. Figure (7) shows the segmentation of the disparity map in two parts. On the left side, points with negative disparities, that is points in front of the reference plane, are shown. The intensity encodes closeness to the camera. Similarly, the right side of the figure shows the points with positive disparities, that is the points which are beyond the reference plane.

5.2 The affine calibration

If we add to the computation of the fundamental matrix, the computation of the homography of the plane at infinity, using the method described in section (3.4), P becomes unnecessary

since its role falls now to the plane at infinity. The only information on the scene to give to the system is a point M_0 at known distance d_0. According to equation (38), M_0 then allows to compute μ, so that, for each other point M,

$$d(M, P'_f) = d_0 \frac{\hat{z}'_0 \mathcal{D}_{\infty 0}}{\hat{z}' \mathcal{D}_\infty}$$

which now makes the robot able, for instance, to stay at a given distance of some points.

Figure (8) shows some real sequences used to perform the affine calibration of a stereoscopic system. Six strong correspondences between the four images have been extracted, from which fifteen correspondences of points at infinity have been computed to finally get the homography of the plane at infinity. Figure (9) shows some midpoints obtained once the system calibrated: the endpoints are represented as dark squares and the midpoints as black crosses.

6 Conclusion

The application of non-metric vision to robot navigation given in this paper shows that some tasks may be performed without requiring full calibration of the system. The metric information, if any, necessary to the task is collected at a well determined stage, using a very simple calibration process.

The affine calibration, which is performed through the process described in this article, seems to offer an interesting framework, in which more sophisticated tasks may be realized, like, for instance, the visual servoing on the midline of a corridor.

Current efforts are devoted to a realtime robust implementation of these ideas, which should be completed at the time of the conference, to the exploration of applications of these ideas to more vision guided tasks and to an investigation of related theoretical problems.

References

[1] Olivier Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[2] Olivier Faugeras, Tuan Luong, and Steven Maybank. Camera self-calibration: theory and experiments. In Giulio Sandini, editor, *2nd European Conference on Computer Vision*, pages 321–334, Santa-Margherita, Italy, May 1992. Springer-Verlag, Lecture Notes in Computer Science 588.

[3] Olivier D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In Giulio Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pages 563–578. Springer-Verlag, Lecture Notes in Computer Science 588, May 1992.

[4] Olivier D. Faugeras and Francis Lustman. Let us suppose that the world is piecewise planar. In O. D. Faugeras and Georges Giralt, editors, *Robotics Research, The Third International Symposium*, pages 33–40. MIT Press, 1986.

[5] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proceedings 4th Alvey Conference*, pages 147–151, Manchester, August 1988.

[6] Richard Hartley, Rajiv Gupta, and Tom Chang. Stereo from Uncalibrated Cameras. In *Proceedings of CVPR92, Champaign, Illinois*, pages 761–764, June 1992.

[7] H. Wang L. S. Shapiro and J. M. Brady. A matching and tracking strategy for independently-moving, non-rigid objects. In *Proceedings of BMVC*, 1992.

[8] Q.-T. Luong, R. Deriche, O.D. Faugeras, and T. Papadopoulo. On determining the Fundamental matrix: analysis of different methods and experimental results. Technical Report RR-1894, INRIA, 1993.

[9] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. Technical Report UCB/CSD-93-772, University of California at Berkeley, Sept 1993.

[10] W. H. Press, B. P. Flannery, S.A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.

[11] L. Robert and O.D. Faugeras. Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair. In *Proc. Fourth International Conference on Computer Vision*, pages 540–544, Berlin, Germany, May 1993.

[12] T. Viéville, C. Zeller, and L. Robert. Recovering motion and structure from a set of planar patches in an uncalibrated image sequence. In *Proceedings of ICPR94*, Jerusalem, Israel, Oct 1994.

[13] Z. Zhang, R. Deriche, Q.-T. Luong, and O. Faugeras. A robust approach to image matching: Recovery of the epipolar geometry. In *Proc. International Symposium of Young Investigators on Information\Computer\Control*, Beijing, China, Feb 1994.

Figure 4: The paper used to define the plane and the left image of the fist taken as an example.

Figure 5: The left and right rectified images of the fist.

Figure 6: The disparity map obtained from the rectified images of figure (5).

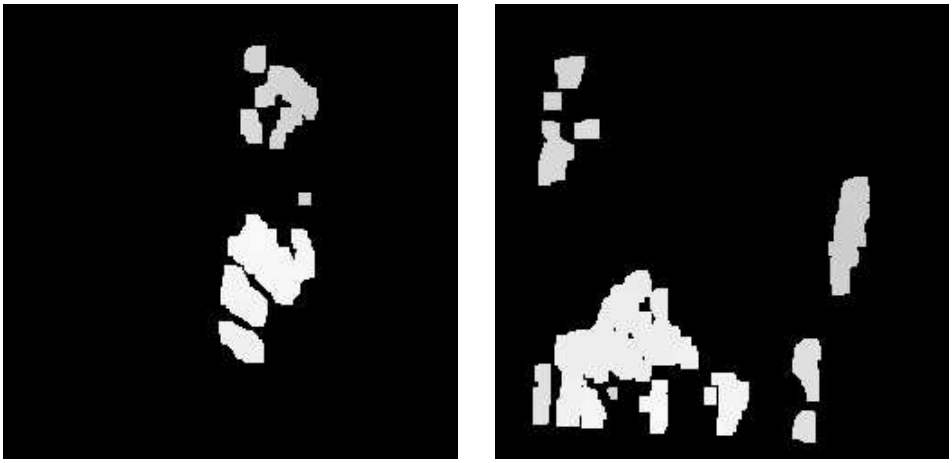

Figure 7: The absolute value of the negative disparities on the left, showing that the fist and a portion of the arm are between the robot and the plane of rectification, and the positive disparities on the right, corresponding to the points located beyond the plane.

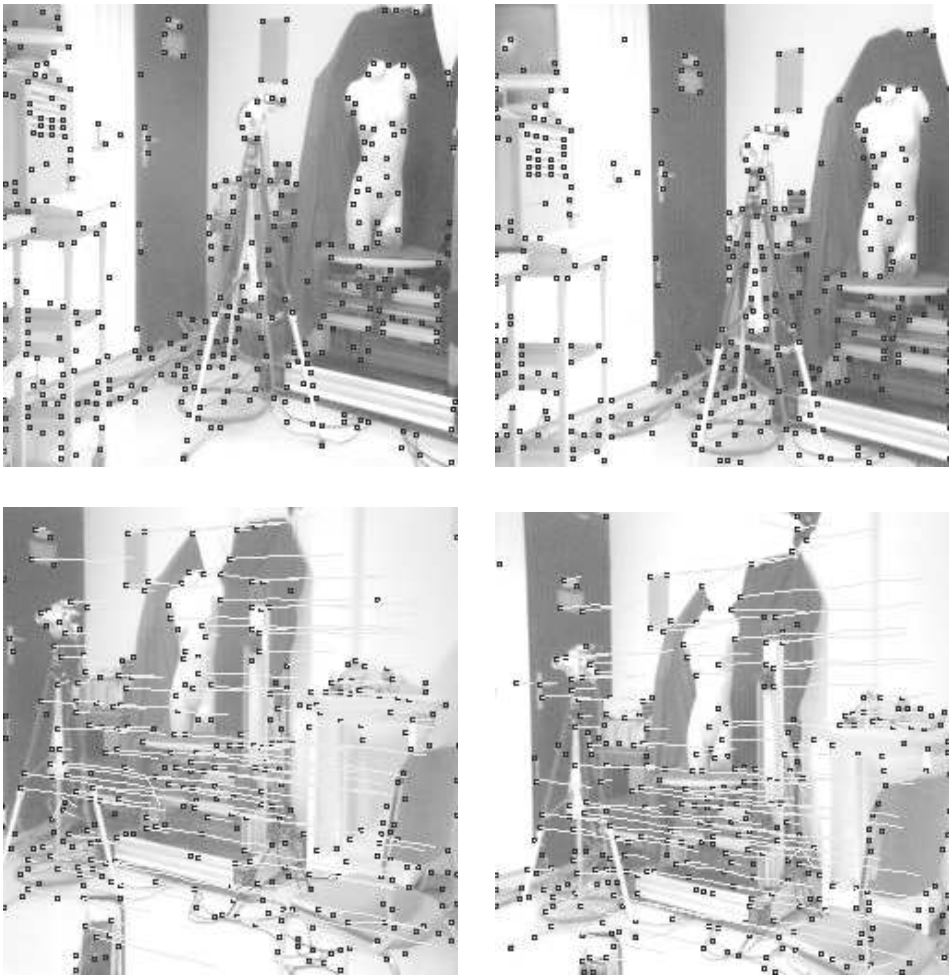

Figure 8: The top images correspond to a first pair of views taken by a stereoscopic system and the bottom images to a second pair taken by exactly the same system after a translation. Among the 297 detected corners of the top left image and the 276 of the top right image, 157 points correspondences have been found by stereo points matching (see section (3.1)), among which 7 outliers have been rejected when computing the fundamental matrix (see section (3.2)). The top to bottom correspondences matching has been obtained by tracking (see section (3.1)).

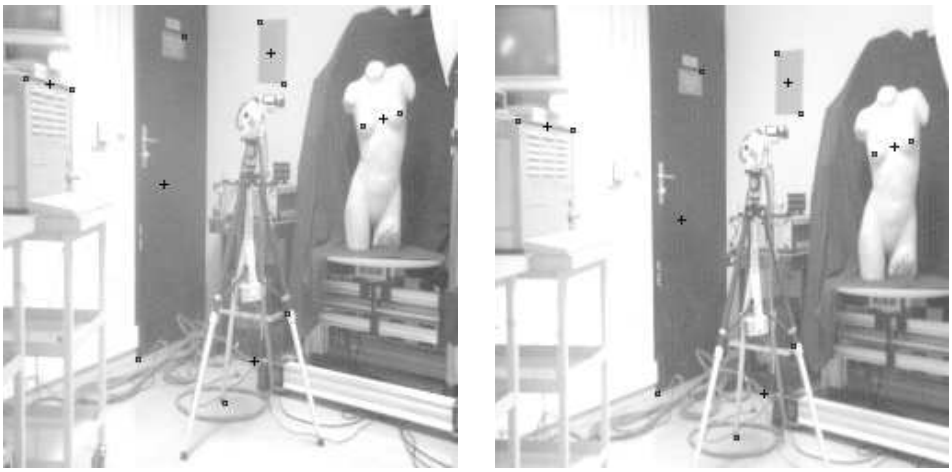

Figure 9: Midpoints obtained after affine calibration (see section (3.4)).

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399